Pseudonorm Approachability and Applications to Regret Minimization

Christoph Dann Yishay Mansour **Mehryar Mohri**
Jon Schneider Balasubramanian Sivan

Google Research

Toulouse Workshop on Learning in Games
June 30, 2024

High-level overview (approachability version)

- Blackwell approachability is a powerful tool for designing learning algorithms (e.g. first algorithms for calibrated forecasts and low internal regret).
- [Abernethy et al., 2011] showed how to reduce approachability to online convex optimization, but this reduction is lossy (e.g., it gets $O(\sqrt{Td})$ regret for full-information learning from d experts when $O(\sqrt{T \log d})$ is possible).
- **This paper**: We show that by using a *pseudonorm* in place of the Euclidean norm in Blackwell approachability, we can recover efficient algorithms with tight regret bounds for many regret minimization problems (external regret, swap regret, and some new ones).

High-level overview (regret version)

Consider a "regret minimization" problem where the learner chooses actions from an n-dimensional convex set, the adversary chooses losses from an m-dimensional convex set, and all payoffs / benchmark payoffs are bilinear in these actions and losses.

Theorem (Informal)

There exists a learning algorithm that guarantees regret $O(\mathrm{poly}(n, m)\sqrt{T})$. Moreover, if it is possible to efficiently compute the regret in hindsight, it is possible to efficiently run this algorithm.

Note that this *does not* depend on the number of benchmarks you are competing against! (e.g., in swap regret there are K^K benchmarks, but $n = m = K$).

Approachability 101

- Let $\mathcal{P} \subseteq \mathbb{R}^n$ and $\mathcal{L} \subseteq \mathbb{R}^m$ be two bounded convex sets.
- Let $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^d$ be any bilinear vector-valued function.
- Let $\mathcal{S} \subseteq \mathbb{R}^d$ be a convex set s.t. for all $\ell \in \mathcal{L}$, there exists a $p \in \mathcal{P}$ such that $u(p, \ell) \in \mathcal{S}$ ($\mathcal{S}$ is *approachable*).

Theorem (Blackwell Approachability)

There exists a learning algorithm (i.e., a function mapping $(\ell_1, \ell_2, \ldots, \ell_{t-1}) \to p_t$) with the property that

$$\lim_{T\to\infty} d\left(\frac{1}{T}\sum_{t=1}^{T} u(p_t, \ell_t), \mathcal{S}\right) = 0,$$

where $d(x, S)$ is the minimum (Euclidean) distance between x and S.

Intuition: multidimensional, algorithmic version of the *minimax theorem*.

Approachability $\implies$ sublinear regret

Full-information **learning with experts** via approachability:

- Let $\mathcal{P} = \Delta_K$, $\mathcal{L} = [0,1]^K$, $\mathcal{S} = (-\infty, 0]^k$, and $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^K$ defined via:

$$u(p, \ell)_i = \langle p, \ell \rangle - \langle e_i, \ell \rangle.$$

- The regret of playing the sequence of actions $\boldsymbol{p} = (p_1, \dots, p_T)$ against the sequence of losses $\boldsymbol{\ell} = (\ell_1, \dots, \ell_T)$ is given by:

$$\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell}) = \left[\max_{i \in [d]} \sum_{t=1}^{T} u(p_t, \ell_t)_i \right]_+$$

- If $\frac{1}{T} \sum_{t=1}^{T} u(p_t, \ell_t) \to \mathcal{S}$, then $\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell}) = o(T)$! *(but what's the dependence on T and K?)*

L_∞-approachability

Note that (when $\mathcal{S} = (-\infty, 0]^d$):

$$\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell}) = \left[\max_{i \in [d]} \sum_{t=1}^{T} u(p_t, \ell_t)_i\right]_+ = T \cdot d_\infty\left(\frac{1}{T}\sum_{t=1}^{T} u(p_t, \ell_t), \mathcal{S}\right). \quad (1)$$

where $d_\infty(x, y) = \max |x_i - y_i|$ is the L_∞ distance.

Two takeaways:

1. Good regret minimization bounds are the same as good (quantitative) bounds for L_∞ Blackwell approachability.
2. We can define $\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell})$ as in (1) for *any* bilinear u where $\mathcal{S} = (-\infty, 0]^d$ is approachable. Captures a fairly wide range of regret minimization problems...

Some other regrets

Swap regret

Let $\mathcal{P} = \Delta_K$, $\mathcal{L} = [0,1]^K$, $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^{(K^K)}$, with

$$u(p,\ell)_\pi = \sum_{i=1}^{K} p_i\ell_i - p_i\ell_{\pi(i)},$$

for each $\pi : [K] \to [K]$. $\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell})$ is **swap regret**, $d = K^K$.

"Procrustean" swap regret

Let $\mathcal{P} = \mathcal{L} = B_K = \{x \in \mathbb{R}^K \mid ||x||_2 \leq 1\}$, $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^{O(K)}$, with

$$u(p,\ell)_Q = \langle p, \ell \rangle - \langle Qp, \ell \rangle,$$

for each $Q \in O(K)$ ("orthogonal group"). $\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell})$ is **Procrustean swap regret** (cf. the orthogonal Procrustes problem), $d = \infty$.

Algorithms for L_∞-approachability

Theorem ([Perchet, 2015, Shimkin, 2016, Kwon, 2021])

For any $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^d$ (where $(-\infty, 0]^d$ is approachable), there exists a learning algorithm that runs in $\mathrm{poly}(n, m, d)$ time per round and guarantees

$$\mathrm{Reg}(\boldsymbol{p}, \ell) = O(\sqrt{T \log d}).$$

Consequences:

- For swap regret, this is a learning algorithm with $O(\sqrt{TK \log K})$ regret that takes time $K^{O(K)}$ per round.
- For Procrustean swap regret, no non-trivial regret guarantee (since $d = \infty$).

New algorithms for L_∞-approachability

Theorem

For any $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^d$ (where $(-\infty, 0]^d$ is approachable), there exists a learning algorithm that guarantees

$$\text{Reg}(\boldsymbol{p}, \ell) = O(nm\sqrt{T}).$$

Furthermore, if $\text{Reg}(\boldsymbol{p}, \ell)$ can be computed in time $\mathrm{poly}(n, m, T)$, then this algorithm can be implemented in time $\mathrm{poly}(n, m)$ per round.

Consequences:

- For swap regret, this is a learning algorithm with $O(K^2\sqrt{T})$ regret that takes time $\mathrm{poly}(K)$ per round.
- For Procrustean swap regret, this is a learning algorithm with $O(d^2\sqrt{T})$ regret that takes time $\mathrm{poly}(\mathrm{d})$ per round.

We can also obtain slightly tighter bounds (in terms of the sizes of $\mathcal{P}$, $\mathcal{L}$, and u) – see paper.

New algorithms for L_∞-approachability (cont.)

But sometimes (e.g., for swap regret) $O(\sqrt{T \log d})$ is better than $O(\mathrm{poly}(n, m)\sqrt{T})$!

Theorem

If there is a $\mathrm{poly}(n, m)$*-time algorithm which, given a (1-dim) bilinear function* $v : \mathcal{P} \times \mathcal{L} \to \mathbb{R}$*, returns the maximum entropy of a distribution* $\rho \in \Delta_d$ *satisfying*

$$\sum_{i=1}^{d} \rho_i u(p, \ell)_i = v(p, \ell),$$

then there exists a learning algorithm that runs in $\mathrm{poly(n, m)}$ *time per round which guarantees* $\mathrm{Reg}(\boldsymbol{p}, \boldsymbol{\ell}) = O(\sqrt{T \log d})$.

This is possible for swap regret, and therefore we can get an $O(\mathrm{poly}(K))$-time algorithm with $O(\sqrt{KT \log K})$ regret.

Online linear optimization

The *online linear optimization (OLO)* problem for (bounded/convex) action set $\mathcal{X} \subseteq \mathbb{R}^k$ and loss set $\mathcal{Y} \subseteq \mathbb{R}^k$:

- Every round $t \in [T]$, the adversary selects a loss $y_t \in \mathcal{Y}$, the learner picks an action $x_t \in \mathcal{X}$ based on y_1 through y_{t-1}.
- The learner wants low regret (wrt best fixed action):

$$\mathrm{Reg}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{t=1}^{T} \langle x, y \rangle - \min_{x^* \in \mathcal{X}} \langle x^*, y \rangle.$$

Algorithms: (all variants of FTRL)

- For general convex $\mathcal{X}, \mathcal{Y}$ can get $O(\mathrm{diam}(\mathcal{X})\mathrm{diam}(\mathcal{Y})\sqrt{T})$ regret ("quadratic regularizer").
- If $\mathcal{X} = \Delta_d$ and $\mathcal{Y} = [-1, 1]^d$, can get $O(\sqrt{T \log d})$ regret ("negentropy regularizer").

Reducing L_∞-approachability to OLO

Initialization: Let $\mathcal{F}$ be an OLO algorithm for the sets $\mathcal{X} = \Delta_d$ and $\mathcal{Y} = [-1, 1]^d$, let x_1 be an arbitrary point in $\mathcal{X}$.
for $t = 1$ to T **do**
    Choose $p_t \in \mathcal{P}$ s.t. $\forall \ell \in \mathcal{L}$, $\langle x_t, u(p_t, \ell) \rangle \leq 0$. *(possible since $(-\infty, 0]^d$ is approachable)*
    Play action p_t and receive as feedback $\ell_t \in \mathcal{L}$.
    Set $y_t = -u(p_t, \ell_t)$.
    Set $x_{t+1} = \mathcal{F}(y_1, y_2, \ldots, y_t)$.
end for

Algorithm 1: Algorithm $\mathcal{A}$ for L_∞-approachability

Theorem ([Perchet, 2015, Shimkin, 2016, Kwon, 2021])

In Algorithm 1, $\text{Reg}_{\mathcal{A}}(\boldsymbol{p}, \boldsymbol{\ell}) \leq \text{Reg}_{\mathcal{F}}(\boldsymbol{x}, \boldsymbol{y})$.

Dimensionality reduction

Observation: A bilinear $u : \mathcal{P} \times \mathcal{L} \to \mathbb{R}^d$ is a linear function from $\mathrm{conv}(\mathcal{P} \otimes \mathcal{L}) \to \mathbb{R}^d$. So:

- $\sum_t u(p_t, \ell_t)$ only depends on $\sum_t p_t \otimes \ell_t$.
- For any $x \in \mathbb{R}^d$, there is a nm-dim $\tilde{x} \in \mathrm{conv}(\mathcal{P} \otimes \mathcal{L})^*$ s.t.

$$\left\langle x, \sum_t u(p_t, \ell_t) \right\rangle = \left\langle \tilde{x}, \sum_t p_t \otimes \ell_t \right\rangle$$

- For any convex set $\mathcal{S} \subseteq \mathbb{R}^d$, there is a convex set $\widetilde{\mathcal{S}} \subseteq \mathrm{conv}(\mathcal{P} \otimes \mathcal{L})$ s.t. $\sum_t u(p_t, \ell_t) \in \mathcal{S}$ iff $\sum_t p_t \otimes \ell_t \in \widetilde{\mathcal{S}}$.
- There exists a pseudonorm f and a convex cone $\widetilde{\mathcal{S}}$ s.t.

$$d_\infty \left(\sum_t u(p_t, \ell_t), (-\infty, 0]^d \right) = d_f \left(\sum_t p_t \otimes \ell_t, \widetilde{\mathcal{S}} \right).$$

Pseudonorms and pseudodistances

Definition

A function $f : V \to \mathbb{R}^+$ on a vector space V is a pseudonorm if:

1. $f(0) = 0$
2. $f(\alpha z) = \alpha f(z)$ for all $\alpha \geq 0$ (positive homogeneity)
3. $f(z + z') \leq f(z) + f(z')$.

A pseudonorm f defines a pseudodistance $d_f(x, y) = f(x - y)$.

Norms $\Leftrightarrow$ symmetric convex sets, pseudonorms $\Leftrightarrow$ convex sets.

Takeaway: Any d-dimensional L_∞-approachability can be reinterpreted as a nm-dimensional pseudonorm-approachability problem.

Reducing L_∞-approachability to OLO (Part II)

Initialization: Let $\mathcal{F}'$ be an OLO algorithm for the sets $\mathcal{X} = \mathrm{conv}(\{u_i\}) \subseteq \mathrm{conv}(\mathcal{P} \otimes \mathcal{L})^*$ and $\mathcal{Y} = -\mathrm{conv}(\mathcal{P} \otimes \mathcal{L})$, let x_1 be an arbitrary point in $\mathcal{X}$.
for $t = 1$ to T **do**
  Choose $p_t \in \mathcal{P}$ s.t. $\forall \ell \in \mathcal{L}$, $\langle x_t, p_t \otimes \ell \rangle \leq 0$.
  Play action p_t and receive as feedback $\ell_t \in \mathcal{L}$.
  Set $y_t = -p_t \otimes \ell_t$.
  Set $x_{t+1} = \mathcal{F}'(y_1, y_2, \ldots, y_t)$.
end for

Algorithm 2: Algorithm $\mathcal{A}'$ for L_∞-approachability (via dimensionality reduction)

Theorem (BMMSS)

In Algorithm 2, $\mathrm{Reg}_{\mathcal{A}'}(\boldsymbol{p}, \boldsymbol{\ell}) \leq \mathrm{Reg}_{\mathcal{F}'}(\boldsymbol{x}, \boldsymbol{y})$.

Key Proof Tool

Theorem

Let f be a pseudonorm and $\mathcal{T}_f^* = \left\{\theta \colon \forall z \in \mathbb{R}^d, \langle \theta, z \rangle \leq f(z)\right\}$ the dual set associated to f. Then, for any closed convex set $\mathcal{S} \subset \mathbb{R}^d$, the following equality holds for any $z \in \mathbb{R}^d$:

$$d_f(z, \mathcal{S}) = \inf_{s \in \mathcal{S}} f(z - s) = \sup_{\theta \in \mathcal{T}_f^*} \left\{ \theta \cdot z - \sup_{s \in \mathcal{S}} \theta \cdot s \right\}.$$

Proof. Using Fenchel-duality:

$$d_f(z, \mathcal{S}) = \inf_{s \in \mathcal{S}} f(z - s) = \inf_{s \in \mathbb{R}^d} \{ f(z - s) + I_{\mathcal{S}}(s) \} \qquad \text{(def. of } I_{\mathcal{S}}\text{)}$$

$$= \sup_{\theta \in \mathbb{R}^d} \left\{ -\left(I_{\mathcal{T}_f^*}(-\theta) + \theta \cdot z \right) - \sup_{s \in \mathcal{S}} \{ -\theta \cdot s \} \right\}$$

($I_{\mathcal{T}_f^*}$ conjugate of f + Fenchel duality theorem)

$$= \sup_{-\theta \in \mathcal{T}_f^*} \left\{ -\theta \cdot z - \sup_{s \in \mathcal{S}} \{ -\theta \cdot s \} \right\} = \sup_{\theta \in \mathcal{T}_f^*} \left\{ \theta \cdot z - \sup_{s \in \mathcal{S}} \theta \cdot s \right\}.$$

More Details

For any $i \in [d]$, there exists $v_i \in \mathbb{R}^{mn}$ such that:

$$\forall p \in \mathcal{P}, l \in \mathcal{L}, \quad u_i(p, l) = \langle v_i, p \otimes l \rangle.$$

Define

$$\widetilde{\mathcal{S}} = \{x \colon \langle x, v_i \rangle \leq 0, \forall i \in [d]\} \quad \text{and} \quad f(x) = \left[\max_{i \in [d]} \langle x, v_i \rangle\right]_+ .$$

$\widetilde{\mathcal{S}}$ is closed convex and f is a pseudonorm. Using the tool result, we can prove:

$$d_\infty \left(\sum_t u(p_t, \ell_t), (-\infty, 0]^d \right) = d_f \left(\sum_t p_t \otimes \ell_t, \widetilde{\mathcal{S}} \right) = \sup_{x \in \mathcal{T}_f^*} \langle x, z \rangle,$$

as well as:

$$\mathcal{T}_f^* = \text{conv}\{v_1, \ldots, v_d\}$$

$$\text{(separability)} \quad \forall \ell \in \mathcal{L}, \exists p \in \mathcal{P} \text{ such that } p \otimes l \in \widetilde{\mathcal{S}}.$$

Proof of Theorem

In light of the previous identities:

$$\begin{aligned}
d_f\left(\frac{1}{T}\sum_{t=1}^{T} p_t \otimes \ell_t, \widetilde{\mathcal{S}}\right) &= \sup_{x \in \mathcal{T}_f^*} \left\langle x, \frac{1}{T}\sum_{t=1}^{T} p_t \otimes \ell_t \right\rangle \\
&= -\inf_{x} \left\langle x, \frac{1}{T}\sum_{t=1}^{T} (-p_t \otimes \ell_t) \right\rangle \\
&= \frac{1}{T}\mathrm{Reg}(\mathcal{F}') - \frac{1}{T}\sum_{t=1}^{T} \langle x_t, -p_t \otimes \ell_t \rangle \\
&\leq \frac{1}{T}\mathrm{Reg}(\mathcal{F}').
\end{aligned}$$

Summary

- If you don't do dimensionality reduction, you have to solve OLO over d-dimensional nice sets (Δ_d and $[-1, 1]^d$).
- If you do dimensionality reduction, you instead can solve OLO over nm-dimensional "weird" sets ($\mathcal{X}$ and $\mathcal{Y}$).
- If you apply generic OLO algorithms (e.g. GD), you get $O(\mathrm{poly}(n, m)\sqrt{T})$ regret.
    - You can do this efficiently with appropriate oracles for $\mathcal{X}$ and $\mathcal{Y}$, which corresponds to evaluating Reg.
- The negentropy regularizer in d dimensions corresponds to the "max entropy" regularizer described earlier (running this gives $O(\sqrt{T \log d})$ regret).

Applications

Two applications (see paper for details):

- **Converging to Bayesian correlated equilibria**: [Mansour et al., 2022] define “Bayesian swap regret”, with the property that if all players in a game are running algorithms with low Bayesian swap regret, then they converge to a Bayesian correlated equilibrium. We provide the first polynomial (time and regret) algorithm for this notion of regret.
- **Reinforcement learning in constrained MDPs**: Existing work (e.g. [Miryoosefi and Jin, 2021]) has applied (L_2) approachability to get low regret algorithms for episodic RL in constrained MDPs. By applying pseudonorm approachability, we improve dependence on number of constraints d from $\mathrm{poly}(d)$ to $\log d$.

Thanks!

Some open questions:

- Is this reduction tight? Can you always recover the *best* approachability guarantees by applying some OLO algorithm?
- When can we efficiently (in $\mathrm{poly}(m, n)$ time) solve the max entropy problem?
- What is the best (efficient) OLO algorithm for the $\mathcal{X}$ and $\mathcal{Y}$ produced by a fixed $\mathcal{P}$, $\mathcal{L}$, and u? Is it an FTRL algorithm?
- What class of regret minimization problems can be captured by (L_∞-approachability)? Other natural applications?

`arxiv.org/abs/2302.01517`

References

Yishay Mansour, Mehryar Mohri, Jon Schneider, and Balasubramanian Sivan.
Strategizing against Learners in Bayesian Games
Conference on Learning Theory (2022).

Sobhan Miryoosefi and Chi Jin.
A Simple Reward-free Approach to Constrained Reinforcement Learning
International Conference on Machine Learning (2021).

Jacob D. Abernethy, Peter L. Bartlett, and Elad Hazan.
Blackwell Approachability and No-Regret Learning are Equivalent
Conference on Learning Theory (2022).

Vianney Perchet.
Exponential weight approachability, applications to calibration and regret minimization.
Dynamic Games and Applications (2015).

Nahum Shimkin.
An online convex optimization approach to Blackwell's approachability.
The Journal of Machine Learning Research (2016).

Joon Kwon.
Refined approachability algorithms and application to regret minimization with global costs.
The Journal of Machine Learning Research (2021).